

Poster II-1

Stratification and Synchronization Inference Technology: Applying Bioinformatics Tools to Clinical Data

Hochberg, Alan M.

ProSanos Corp., Harrisburg, PA, USA

In biomedical research, it is increasingly important to correlate clinical information with genomic data such as single-nucleotide polymorphisms and gene expression profiles. Although much attention has been given to the analysis of genomic data itself, less work has been focused on the problems of preparing and classifying clinical data from patient medical records, in a way that is specifically tailored to biomedical discovery.

We describe a computational tool called Stratification and Synchronization Inference Technology (SSIFT), which provides a new way to classify patient populations, based on measurements that describe the evolution of a each patient's disease over time. The groupings identified by SSIFT are well suited to correlation analysis with genomic data, or with treatment outcomes in pharmaceutical trials.

SSIFT accounts for the fact that patient records are often fragmentary views of a long-term disease process. Data collection begins when a patient presents symptoms to a physician, not when the patient's disease process actually starts, and typically lasts for only a portion of the disease duration. To generate a model of the entire disease process, SSIFT begins with curve fitting, and then employs a "fragment assembly" approach, similar to that used in gene bioinformatics, to synchronize, or align in time, clinical data records. The synchronized data can then be clustered, using conventional clustering algorithms. The software assigns each patient to one of a set of disease progression patterns, or trajectories, that a disease can follow over time, along with an indication as to where in time that patient fits along the trajectory.

The most optimal application for SSIFT is in complex, multi-factorial diseases, which have many comorbid conditions, a long, variable time course, and variations from patient to patient in the degree of treatment-responsiveness. An example of SSIFT analysis of diabetes data will be shown. This approach to clinical data analysis applies equally to basic research, drug discovery, clinical trial design, and pharmaceutical outcomes research.